Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Modeling Students' Ability to Recognize and Review Graded Answers that Require Immediate Attention

Yancy Vance PAREDES[a*] & I-Han HSIAO[b]
[a]*Arizona State University, USA*
[b]*Santa Clara University, USA*
*yvmparedes@asu.edu

Abstract: Students utilize various resources to prepare for an examination, such as lecture materials, homework, or previous quizzes or tests. Reviewing graded tests allows students to develop their metacognitive skills. However, a lack of proper guidance, exacerbated by a lack of maturity, hinders fully realizing the benefits of learning from past mistakes. In this paper, we investigated students' reviewing strategies. We analyzed the clickstream data of students taking a Computer Science Education course. Using Hidden Markov models (HMMs), we modeled the reviewing behaviors of high-performing and low-performing students. Our preliminary findings suggest that the two groups share some similar strategies but also have some that are particular to the group.

Keywords: Reviewing behavior, clickstream data, hidden Markov models

1. Introduction

Students prepare for examinations using various resources that are made available to them. In an earlier survey, students indicated that apart from lecture materials, they also reexamine previous quizzes or tests (Paredes et al., 2017). Students used it as a practice opportunity anticipating that a similar question would come out in the examination. Reviewing assessments enables students to demonstrate and enhance their metacognitive skills, such as monitoring mistakes or evaluating a learning strategy's success and adjusting if necessary. Knowing how they performed in a graded assessment allows them to formulate a plan to address their misconceptions.

This paper aims to determine whether students can identify the questions that require their immediate attention. *Do students review questions based on their performance? What reviewing patterns can be uncovered?* These questions can be answered by looking at how students interact with an educational technology that captures their reviewing behaviors. These strategies are captured in the form of clickstream data. Many approaches can be employed to model and interpret such behaviors. In this paper, we modeled students' clickstream behaviors using Hidden Markov models (HMMs) and presented our preliminary findings.

2. Related Work

Earlier works have examined the distribution of the students' review actions and how this affects their succeeding examination performance (Paredes, Azcona, et al., 2018; Paredes et al., 2019). Moreover, when students review their graded tests, they benefit from being guided in identifying which items to focus on (Paredes, Hsiao, et al., 2018). However, these earlier investigations did not consider the dataset's sequential and temporal dimensions. The analyses focused only on the frequency of user actions and did not account for the transitions between them. Therefore, this current work aims to address the said limitation.

HMM is among the popular approaches to analyzing and modeling clickstream data (Rabiner, 1989). Beyond the educational data mining domain, many works have leveraged this technique to understand behavioral patterns (e.g., common transitions as visitors navigate an e-commerce website; Liu et al., 2017). An advantage of this approach is that it incorporates the temporal information of the data as opposed to simple clustering (Perera et al., 2009).

3. Methods

3.1 Data Collection

We analyzed a total of 88,111 actions from clickstream data of 317 students enrolled in an Object-Oriented Programming and Data Structures class. These interactions were captured using the educational tool WebPGA (Paredes et al., 2019). The course had a total of 17 paper-based assessments. Three of them were examinations, while the other 14 were quizzes. Two of the quizzes were for credit, while the rest were not. Students had to answer these quizzes and were awarded full points regardless of the correctness of their answers, as these were used for attendance.

Although the system can capture multiple student interactions, we limited this preliminary analysis to three specific actions. These actions represent the three levels of how a student can review an assessment as illustrated in Figure 1. The first level is the *dashboard* or *class overview* (Figure 1a), where students are presented with a list of all the assessments administered in class and the scores they obtained. The second level is the *assessment overview* (Figure 1b), where students are shown all the questions from the selected assessment. Their scores for the individual questions are shown at this level. From here, students can choose a question to review, which leads them to the third level or the *question overview* (Figure 1c). Students can see fine-grained information about the question in the third level, such as the rubrics used to assess their answer, detailed feedback from the grader on why such a score was given, and written annotations on the digital paper.

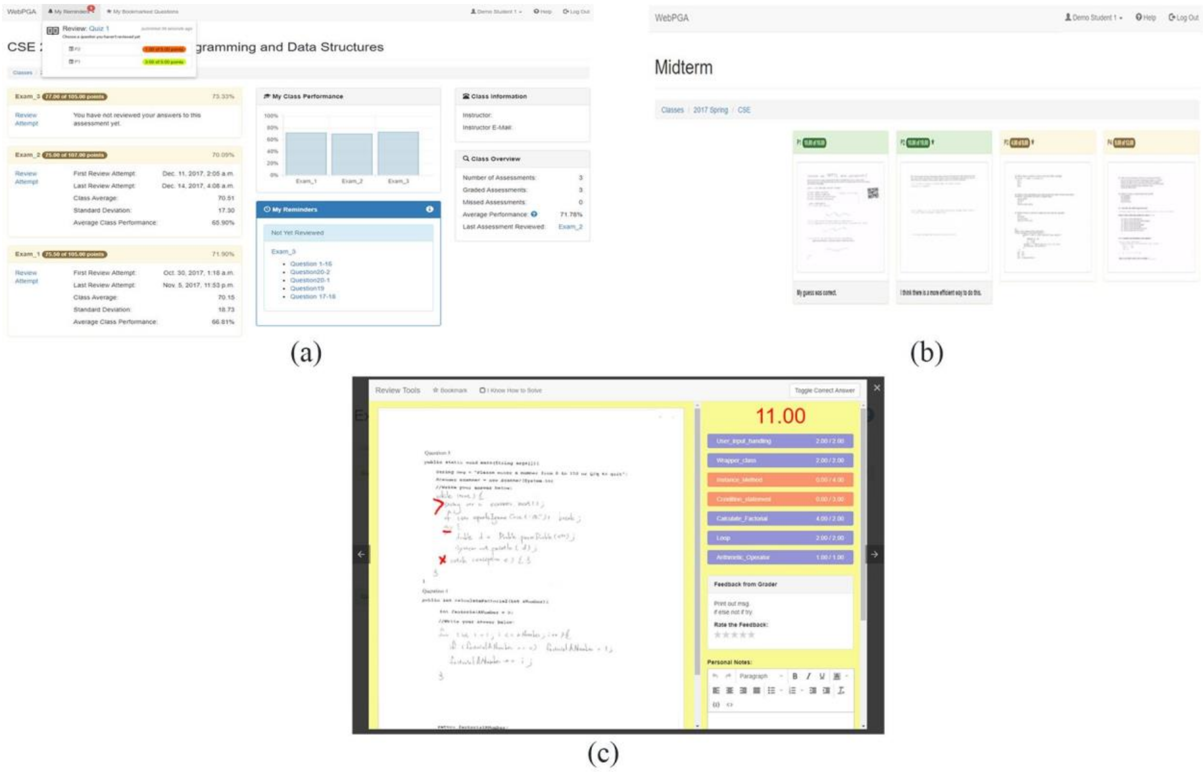

Figure 1. Screenshot of the various levels how a student can review an assessment: (a) dashboard or class overview, (b) assessment overview, and (c) question overview

3.2 Data Pre-Processing

The students used the platform throughout the semester at their convenience. They were informed via announcements in the learning management system (i.e., Blackboard) immediately after an assessment was graded. Each student's overall performance was computed by averaging the student's scores in the three examinations. Lastly, students were classified as *high-performing* or *low-performing* using the class average (M=0.83, SD=0.12) as the cut-off point.

Each question had varying difficulty. To determine this, we examined how the entire class performed. The average score for all the answers to a question was computed. The higher the value, the easier the question is. We used this information to add context to the reviewing behavior of the students. We compared the score obtained by the student to the question's difficulty. If the student obtained a higher score, a review action on this graded answer was labeled a *non-urgent question review*; otherwise, it was labeled an *urgent question review*. The rationale behind this heuristic is that students should attend to the questions they did not satisfactorily do the soonest.

Each student is represented by a single sequence enumerating the various actions performed on the system. The system can identify a group of actions performed in a single session. Therefore, a symbol was introduced to indicate the beginning of a new session. The average sequence length was 125 actions. Table 1 provides a summary of the various symbols used for the analysis.

Table 1. *Symbols used to represent the various actions performed by the students*

Symbol	Description
D	Viewing the class dashboard that shows an overview of the student’s scores on all the assessments.
A	List all the questions of an assessment and the scores obtained by the student in each question. Allows them to choose a question to review in detail.
N	Reviewing a graded answer considered non-urgent. The student's score is above the threshold based on the question's difficulty.
U	Reviewing a graded answer considered urgent. The student's score is below the threshold based on the question's difficulty.
X	Reviewing an ungraded question. The question's difficulty is unknown.
S	Used as a marker to denote the beginning of another session in the student's sequence.

3.3 Hidden Markov Model

One common approach to modeling sequential data is through HMM (Rabiner, 1989). We developed two HMMs to model the sequences of the two student groups, one for each group, and explore any similarities. The number of hidden states (HS) was a parameter that needed to be estimated. The parameter was set to four based on a similar early work where the Akaike Information Criterion (AIC) was used to determine the optimal number of hidden states (Hsiao et al., 2017). As shown in Table 2, each HS represents a strategy where the emission probabilities of each action are identified. The most probable action of a strategy is highlighted. Essentially, an HS encapsulates the combination of actions that are likely to be done by the student. The transition probabilities between strategies (HS) of the two models are illustrated in Figure 2. Due to the system's design, the prior probability of the HS1 for both models is 1.00. It simply means that all sequences always begin with navigation from the dashboard.

Table 2. *The Emission Probabilities of the Two HMMs*

Group	Strategy	D	A	N	U	X	S
High	HS1	**0.70**	-	-	-	-	0.29
	HS2	-	**0.95**	-	-	0.03	0.03
	HS3	0.21	0.06	0.21	0.16	**0.36**	
	HS4	-	0.15	**0.63**	0.22	-	-
Low	HS1	**0.59**	0.12	0.01	-	0.28	-
	HS2	-	**0.79**	0.09	0.12	-	-
	HS3	0.09	**0.60**	-	-	0.04	0.27
	HS4	0.03	-	0.31	**0.66**	-	-

Note. Values less than 0.01 were omitted.

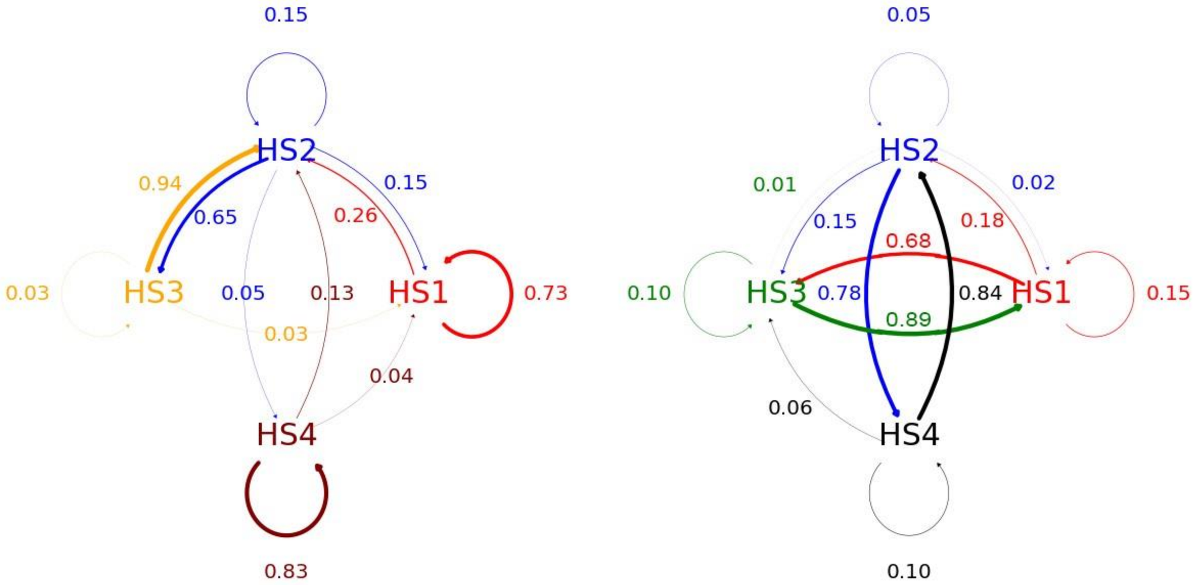

Figure 2. The transition probabilities of the two HMMs: high-performing students (left) and low-performing students (right), both with four hidden states (HS), each representing a strategy

4. Preliminary Findings

Based on how the system was design, a typical workflow always begins with the class dashboard (Figure 1a), where students choose an assessment to review. Afterward, students choose a particular question to review (Figure 1b). Students can navigate to the next question using the next or previous buttons; or close the current question window, go back to the assessment overview, and choose another question from the list (Figure 1c). The system has a personalized notification panel on the top and is always visible regardless of where the student is. It enables students to navigate directly to specific questions not yet reviewed.

The hidden states reflect the students' reviewing strategies. As evident in the two groups' prior probabilities, students would always start their review from the dashboard. However, the high-performing group's emission probability for the HS1 strategy is limited only to actions D and S. This means they do not go into the details of any assessments at that moment. On the other hand, the HS1 strategy of the low-performing group shows more actions aside from the dashboard. They would go ahead and review ungraded questions. This action by the low-performing group could indicate that they were using the notification panel.

High-performing students are likely to repeat their HS1 strategy of checking their overall performance from time to time. However, there are instances where they would change strategy and go into the details of an assessment, then later details of various questions as evidenced by their transition from HS1➔HS2➔HS3. Interestingly, in their HS3 strategy, the emission probability is high on ungraded questions, which suggests that they exert effort to review questions that were not graded to help them prepare for an exam. The loop in HS3➜HS2➜HS3 indicates that these students consciously determine which questions to review next instead of simply relying on the built-in navigation buttons. It possibly suggests the ability of these students to recognize which of their graded answers to review next. This strategy could be a potential indicator of the student's awareness of planning on how to address their misconceptions. The HS4 strategy, which focused on reviewing the non-urgent questions, had a lower likelihood of happening since the only way to reach HS4 is through itself or from HS2.

Low-performing students, like the high-performing students, had a high probability of transitioning to a strategy involving the assessment overview, HS2 or HS3 (more probable). A closer look into the more probable transition HS3 strategy's emission probability shows the presence of seeing the session marker. It suggests that these students often stopped reviewing at the assessment level and did not proceed further to the question level. It is even more pronounced in the following transition of HS3➜HS1➜HS3, meaning they would only log in to the system to check their scores without the intention of knowing where they made mistakes or learning from the feedback provided by the grader. The strategy for reviewing questions, particularly urgent ones, in detail HS2➜HS4➜HS2 involves a loop. These states can only be reached from the HS1 strategy.

5. Limitations and Future Work

This preliminary work aimed to model students' ability to recognize questions requiring immediate attention as they review and prepare for an upcoming examination. It also investigated whether the two student groups had different strategies in this process. One of the limitations of this analysis involves estimating the parameters of the HMM. Although we followed the AIC method, several approaches can be explored that use information from the data. For example, Li and Biswas (2002) proposed a Bayesian approach to estimate the number of hidden layers based on the data.

Two other approaches to analyzing sequential data include clustering students who had a similar distribution of actions they performed. Another is leveraging sequential pattern algorithms (e.g., Generalized Sequential Pattern; Srikant & Agrawal, 1996) to identify frequently performed actions. Differential pattern mining (Kinnebrew et al., 2013) which focuses on sequences specific only to certain student groups, is also a promising direction. Finally, instead of focusing on what actions are frequently performed on the system, another perspective is to examine each student group's distinct actions.

The clickstream data used in this study focused only on what was available on the system. This data can be used to complement other clickstream data available from other systems, such as learning management systems. In effect, it would allow for a better understanding of the students, as shown in the work of Gitinabard and colleagues (2019). With the shift of most activities to online due to the Covid-19 pandemic, it would also be worth investigating whether similar trends can be found in assessments administered electronically.

The current models can be incorporated into the system, allowing future studies to investigate whether students would benefit from a personalized intervention to improve their reviewing behaviors. By analyzing the students' clickstream data in real-time, tailored suggestions in the form of notifications can be shown to students, making them aware of their current strategy. The same can also be used to make them understand the strategies of successful students, hopefully enabling other students to emulate such behaviors.

References

Gitinabard, N., Barnes, T., Heckman, S., & Lynch, C. F. (2019). What will you do next? A sequence analysis on the student transitions between online platforms in blended courses. In *Proceedings of the 12th International Conference on Educational Data Mining* (pp. 59–68).

Hsiao, I.-H., Huang, P.-K., & Murphy, H. (2017). Uncovering reviewing and reflecting behaviors from paper-based formal assessment. In *Proceedings of the 7th International Learning Analytics & Knowledge Conference* (pp. 319–328).

Kinnebrew, J. S., Loretz, K. M., & Biswas, G. (2013). A contextualized, differential sequence mining method to derive students' learning behavior patterns. *Journal of Educational Data Mining, 5*(1), 190–219.

Li, C., & Biswas, G. (2002). A bayesian approach for structural learning with hidden markov models. *Scientific Programming, 10*(3), 201–219.

Liu, Z., Wang, Y., Dontcheva, M., Hoffman, M., Walker, S., & Wilson, A. (2017). Patterns and sequences: Interactive exploration of clickstreams to understand common visitor paths. *IEEE Transactions on Visualization and Computer Graphics, 23*(1), 321–330.

Paredes, Y. V., Azcona, D., Hsiao, I.-H., & Smeaton, A. F. (2018). Learning by reviewing paper-based programming assessments. In *Proceedings of the 13th European Conference on Technology Enhanced Learning* (pp. 510–523).

Paredes, Y. V., Hsiao, I.-H., & Lin, Y. (2018). Personalized guidance on how to review paper-based assessments. In *Proceedings of the 26th International Conference on Computers in Education* (pp. 257–265).

Paredes, Y. V., Huang, P.-K., & Hsiao, I.-H. (2019). Utilising behavioural analytics in a blended programming learning environment. *New Review of Hypermedia and Multimedia, 25*(3), 89–111.

Paredes, Y. V., Huang, P.-K., Murphy, H., & Hsiao, I.-H. (2017). A subjective evaluation of web-based programming grading assistant: Harnessing digital footprints from paper-based assessments. In *Proceedings of the 6th Multimodal Learning Analytics Workshop and the 2nd Cross-LAK Workshop* (pp. 23–30).

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaane, O. R. (2009). Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering, 21*(6), 759–772.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE, 77*(2), 257–286.

Srikant, R., & Agrawal, R. (1996). Mining sequential patterns: Generalizations and performance improvements. In *Proceedings of the 5th International Conference on Extending Database Technology* (pp. 1–17).